IN THE UNITED STATES DISTRICT COURT
FOR THE MIDDLE DISTRICT OF TENNESSEE
AT NASHVILLE

FRED ROBINSON; ASHLEY SPRAGUE; JOHNNY GIBBS; and BRIANNA BOOHER, on behalf of themselves and all others similarly situated, Plaintiffs, v. DAVID W. PURKEY, Commissioner of the Tennessee Department of Safety and Homeland Security, in his official capacity; DEBBIE MOSS, Circuit Court Clerk of Wilson County, Tennessee, in her official capacity; MELISSA HARRELL, Circuit Court Clerk of Rutherford County, Tennessee, in her official capacity; COREY LINVILLE, Court Clerk of the Municipal Court of Lebanon, Tennessee, in his official capacity; SUSAN GASKILL, Court Clerk of the City Court of Mt. Juliet, Tennessee, in her official capacity; WILSON COUNTY, TENNESSEE; RUTHERFORD COUNTY, TENNESSEE; LEBANON, TENNESSEE; and MT. JULIET, TENNESSEE, Defendants.	)	No. 3:17-cv-01263 **JUDGE TRAUGER**

DECLARATION OF PROF. DAIN DONELSON IN SUPPORT OF PLAINTIFFS' MOTION TO CERTIFY SUBCLASSES

Dain Donelson declares under penalty of perjury pursuant to 28 U.S.C. § 1746 as follows:

1. I am an Associate Professor and KPMG Faculty Fellow at the McCombs School of Business at the University of Texas at Austin, and I hold a courtesy appointment as an Associate Professor in the University of Texas School of Law. I received my Ph.D. in Accountancy from the University of Illinois in 2007 and my J.D. from Northwestern University School of Law in 1997. A copy of my *Curriculum Vitae* is annexed hereto as Exhibit 1.

2. My principal area of research is the use of empirical and statistical methods to evaluate the effect of legal rules on real-world behavior. I focus primarily on legal rules relating to accounting and finance, but the empirical and statistical methods I use are general and can be used to analyze the effect of legal rules in other areas as well. I am, among other things, a reviewer for the Conference on Empirical Legal Studies and the *Journal of Empirical Legal Studies*.

3. I have been asked by Plaintiffs in this action whether certain data they have obtained, which are described in the Declaration of Edward P. Krugman dated September 18, 2017 (ECF #19) and the Second Declaration of Edward P. Krugman, which I understand is being filed herewith, permit one to estimate how many of the suspensions of driver's licenses pursuant to Tenn. Code Ann. §§ 55-50-502(a)(1)(H), (I) arising from Wilson County, Tennessee arise from the group of people that the United States Census Bureau records as having incomes below the federal poverty level. Subject to certain qualifications, which I describe below, I have concluded that the data do permit such estimates to be made. The remainder of this Declaration sets forth the bases for that conclusion and the estimates and the qualifications thereto. It also discusses certain implications of the data with respect to suspensions that were not subsequently reinstated.

The Dataset

4. Counsel for Plaintiffs provided me, in electronic form (an Excel workbook), with data concerning the number of suspensions of residents of each of Tennessee's 95 counties pursuant to the Tennessee statutes at issue in this action, which I understand to be Tenn. Code Ann. §§ 55-50-502(a)(1)(H), (a)(1)(I). The workbook also contained data on the number of suspensions that were subsequently reinstated and on the number of individuals in each county who were at or below the federal poverty level.

5. I have not conducted any independent investigation of the source or validity of the data I have been provided, and the discussion and conclusions in this Declaration are based

solely on the data as provided to me. The only change I made to the data was rename the variables for ease of use within the software.

6. The data provided to me were presented separately for Whites and for African-Americans and were limited to those two categories. I understand that Mr. Krugman set forth the reasons for this limitation in his first Declaration as follows:

> I did not obtain poverty data for other races or ethnic groupings, both because the Tennessee suspension dataset does not characterize Hispanics in the same way the Census Bureau does and because Whites and African-Americans together constitute the great bulk both of the population of Tennessee (96%) and the number of suspensions (91%). (ECF #19 at ¶ 13)

As set forth in Paragraph 10 of Mr. Krugman's first Declaration, the data provided to me were limited to suspensions (and reinstatements) characterized as "Failure to Pay ('FTP')" and "Failure to Pay or Appear ('FTP/A')" and did not include suspensions or reinstatements characterized as "Failure to Appear ('FTA')."

7. A printout of the dataset as provided to me is annexed hereto as Exhibit 2.

8. At the time I performed the analysis described in this Declaration, I understood that individuals could have multiple suspensions and that the dataset I had been provided reflected counts of suspensions and reinstatements, not counts of individual drivers suspended or reinstated. I now understand that, in a conversation on August 7, 2018, counsel for Defendant Purkey informed Mr. Krugman that the counts in the dataset were of individuals, not suspensions, and that deposition discovery on that subject is ongoing. The conclusions stated in the next portion of this Declaration are framed in terms of the dataset as I originally understood it—*i.e.*, as counts of suspensions. I have, however, reviewed the analysis on the assumption that the dataset represented counts of individuals instead, and I have concluded that the analysis remains unchanged, although the verbal formulation of the results would change to reflect the change in assumption. I address this further in the last section of this Declaration.

Regression Analysis

9. The tool I used to analyze the data was multiple linear regression, which I applied through the Stata® software package. Multiple linear regression is a statistical technique that uses explanatory variables to attempt to predict the outcome of a response variable. The results of performing a regression analysis are:

- an equation that estimates the relationship between the explanatory variables and the response variable, with each explanatory variable having a "coefficient" that describes the amount of change in the response variable one can expect from a given change in that explanatory variable;
- "confidence intervals" as to each of the coefficients,[1]
- measures of the "statistical significance" of the various coefficients,[2] and
- measures of the extent to which there are likely to be factors influencing the response variable other than the specific explanatory variables chosen for analysis.

10. Because I was asked to estimate the relationship between suspensions and poverty, I chose "Total Suspensions" of residents of each of Tennessee's 95 Counties as the response variable in the regression analysis and chose three explanatory variables whose impact on suspensions would be estimated: the number of Whites in poverty in the County, the number of African-Americans in poverty in the County, and the number of Whites and African-Americans (combined)

---

[1] Regression coefficients are *estimates*. The "confidence interval" for a coefficient is the range within which one can have a specified degree of comfort that the true value of the coefficient lies. The confidence interval most often used is 95%; if one estimates a coefficient as, say, 0.8, with a 95% confidence interval of [0.5, 1.1], that is saying that 19 times out of 20 the true value of the coefficient is somewhere between 0.5 and 1.1.

[2] "Statistical significance" is the flip side of the confidence interval and is a reflection of the likelihood that the true value of the coefficient is not zero—which matters because a zero coefficient to an explanatory variable means that that variable does not in fact influence the response variable. Saying that an explanatory variable is statistically significant at the .05 level is the same as saying that zero is not in the 95% confidence interval for the coefficient (*i.e.*, that there is at most a 5% chance that the true value of the coefficient is zero).

Statistical significance should not be confused with economic (or "real world") significance. It is entirely possible for a variable to be statistically significant (very likely to have an effect on the response variable) while being economically *insignificant* (the likely effect is very small).

not in poverty in the County. I chose separate variables for Whites and African Americans in poverty because I understood from Mr. Krugman (and my inspection of the dataset confirmed) that Whites and African Americans are suspended at different rates in Tennessee.

The Impact of Poverty on the Number of Suspensions

11. The Stata® printout reflecting the regression I performed is annexed hereto as Exhibit 3. The basic equation (with variables renamed for convenience) is as follows:

$$Susp = 0.358514 * AAPov + 0.0643328 * WPov + 0.0183647 * NonPov$$

This means that adding one poor African-American to the population of a County can be expected to produce 0.358514 additional suspensions of residents of that County; adding one poor White person can be expected to produce 0.0643328 additional suspensions; and adding someone who is not poor (White or African American) can be expected to produce 0.0183647 additional suspensions.

12. All three coefficients are significant at the .05 level, and the two poverty coefficients are significant at any level one would reasonably measure. This means that the effects described in the regression equation are very likely to be real-world phenomena and not the results of random chance.

13. The real-world phenomenon described by the equation is that, in terms of the impact of adding one more person to the County's population on the number of suspensions of residents of that County, it matters a great deal whether the person one is adding is poor. Adding a poor White person to the County's population has three-and-a-half times the effect on the number of suspensions as adding a non-poor individual does (0.0643328/0.0183647 = 3.503). Adding a poor African-American person to the population has nearly twenty times that effect (0.358514/0.0183647 = 19.522).

14. In addition, the three explanatory variables explain virtually all of the variation in the number of suspensions in a given County. The accepted measure of explanatory power in a regression is known as "Adjusted R-Squared," and Adjusted R-Squared for this regression is 0.9857. That means that our three explanatory variables explain over 98% of the variation in number of suspensions; the aggregate of all other potential causative factors explains less than 1.5% of that variation. In my own work, I only recall *ever* seeing Adjusted R-Squared this high once, and it was in the context of a relationship between revenues and expenses, which is known *a priori* to be a mechanical one.

Estimating the Likelihood That a Suspension in Wilson County Arises from a Poor Person

15. Using the regression equation estimated above, it is possible to go a step further and estimate the likelihood that a given suspension of a Wilson County resident was of a poor person.

16. The regression model predicts 3,323 suspensions of Wilson County residents. By re-running the model with, first, the number of non-poor people set to zero and, then, the number of poor people set to zero, one can disaggregate the predicted number of suspensions into poor and non-poor, giving an expected number of Wilson County suspensions relating to poor people of 1,415 and an expected number of suspensions relating to non-poor people of 1,908. These "linear predictions" are the calculations numbered "4." and "5." at pages 4-8 of Exhibit 3. Thus, the basic "point estimate" of the proportion of Wilson County suspensions that relate to poor people is 42.6% (*i.e.*, 1,415/3,323).

17. It is also possible to estimate the 95% confidence interval around this 42.6% point estimate, using basic statistical techniques for estimating the margin of error of a proportion

where the numerator (here, poor-people suspensions) is a subset of the denominator (here, all suspensions) and the margins of error of the numerator and denominator are known. Doing that calculation (which, together with a summary of results, is set forth in Exhibit 4 hereto) produces a 95% confidence interval for the proportion of [27.0%, 58.2%]. That is, one can say with 95% confidence that the true likelihood that a given suspension of a Wilson County resident will have arisen from a poor person is somewhere between 27.0% and 58.2%.

Unreinstated Suspensions

18. I have also been asked whether it is possible to estimate the number of unreinstated suspensions (or, more precisely, the number of suspensions less the number of reinstatements) in Tennessee that arise from poor people. Using a similar regression to the one described above, it is possible to make such an estimate, and the answer is that virtually all unreinstated suspensions relate to poor people.

19. Performing a regression with the response variable "Unreinstated Suspensions" and the same explanatory variables as before produces coefficients on the poverty variables that are statistically significant at the .05 level, as before (and Adjusted R-Squared remains very high, at over 93%), but here the non-poverty variable is *not* statistically significant. (The Stata® printout is annexed hereto as Exhibit 5.[3]) That means that one cannot say with a reasonable degree of assurance that "number of persons not in poverty" in a County has *any* impact on the number of unreinstated suspensions arising from that County.

[3] The measure of statistical significance is the "p" statistic; if $p < 0.05$ the result is "significant at the .05 level"; if not, not. The "p" statistic for the non-poverty variable in this regression is $p = 0.224$, and the variable is therefore not significant. Put differently, and as reflected in the printout, zero *is* in the 95% confidence interval for this coefficient, which is given as [-0.0041478, +0.0174927].

20. Saying that one cannot be sure that X has an impact on Y is different from saying that one *can* be sure that X does *not* have an impact on Y. To see if the stronger statement could be made here, I re-ran the regression without the not-in-poverty variable—*i.e.*, with just two variables, number of Whites in poverty and number of African-Americans in poverty. The Stata® printout is annexed hereto as Exhibit 6.

21. The results of running the second regression are that both poverty variables are significant at any conceivable level, and Adjusted R-Squared is virtually unchanged from the first regression.[4] The fact that Adjusted R-Squared of the two poverty variables taken alone is so high, combined with the lack of change when the non-poverty variable is added, permits us to say that any impact of the non-poverty variable on the number of unreinstated suspensions in a given County is confined to the 6.8% of the variation that is not explained by the two poverty variables. Accordingly, one can say with a high degree of confidence that over 93% of unreinstated suspensions in Tennessee relate to poor people.

"Individuals Suspended" vs. "Suspensions"

22. As noted above, it may be the case that the dataset I was provided included counts of individuals suspended and reinstated, not counts of suspensions and reinstatements (individuals can have multiple suspensions). That change in information does not affect the analysis set forth above from a numerical or statistical standpoint: that is, all of the explanatory variables are the same as before, and the regressions and other estimates remain valid as statements concerning the relationships between the explanatory variables and the response variables.

23. In particular, my statements concerning statistical significance, explanatory power, and confidence intervals remain unchanged by the change in assumption.

---

[4] Adjusted R-Squared goes down from 0.9319 to 0.9315, a change of .0004. In other words, Adjusted R-Squared for the two regressions is the same to within 3 decimal places.

24. What does change is the characterization of the response variables and, thus, the verbal formulation of the results. That is, because the response variables would now be "Numbers of Individuals Suspended" and "Unreinstated Individuals"[5] instead of "Numbers of Suspensions" and "Unreinstated Suspensions," the description of the results would differ somewhat. I present here, in bullet point form, the results set forth above with the appropriate recharacterization of the response variables:

- Adding one poor African-American to the population of a County can be expected to produce 0.358514 additional residents of that County suspended; adding one poor White resident can be expected to produce 0.0643328 additional residents suspended, and adding someone who is not poor can be expected to produce 0.0183647 additional residents suspended. (¶ 11)
- In terms of the impact of adding one more person to the County's population on the number of residents suspended, it matters a great deal whether the person one is adding is poor. (¶ 13)
    - Adding a poor White person to the County's population has three-and-a-half times the effect on the number of residents suspended as adding a non-poor individual does. (*Id*.)
    - Adding a poor African-American person to the population has nearly twenty times that effect. (*Id*.)
- The three explanatory variables explain over 98% of the variation in number of residents suspended; the aggregate of all other potential causative factors explains less than 1.5% of that variation. (¶ 14)
- The disaggregation in paragraph 16 is, under the new assumption, into an expected number of poor Wilson County residents having suspensions as 1,415, and the expected number of non-poor residents having suspensions as 1,908, giving a point estimate that 42.6% of suspended Wilson County residents are poor. (¶ 16)
- The 95% confidence interval is the same, so that one can now say with 95% confidence that the true likelihood that a suspended Wilson County resident will be poor is somewhere between 27.0% and 58.2%. (¶ 17)
- Virtually all persons in Tennessee with unreinstated suspensions are poor. (¶ 18)
- Any impact of the non-poverty variable on the number of suspended residents of a given County who do not get reinstated is confined to the 6.8% of the variation that is not explained by the two poverty variables. (¶ 21)

[5] Or, more precisely, the number of suspended individuals less the number of reinstated individuals.

- One can say with a high degree of confidence that over 93% of Tennesseans with unreinstated suspensions are poor.

This day of August, 2018, I declare under penalty of perjury that the foregoing is true to the best of my knowledge and belief.

Dain Donelson

- One can say with a high degree of confidence that over 93% of Tennesseans with unreinstated suspensions are poor.

This 8th day of August, 2018, I declare under penalty of perjury that the foregoing is true to the best of my knowledge and belief.

Dain Donelson